



EXHIBIT A



Calculus Applied to Probability and Statistics

by

Stefan Waner and Steven R. Costenoble

This Section: 3. Mean, Median, Variance and Standard Deviation

2. Probability Density Functions:		4. You're the Expert -	Calculus and	"Real
② Uniform, Exponential, Normal, and Beta	② Section 3 Exercises	③ Expert - Creating a Family Trust	④ Probability Main Page	⑤ "Real World" Page

3. Mean, Median, Variance and Standard Deviation

Mean

In the last section we saw that if saving and loan institutions are continuously failing at a rate of 5% per year, then the associated probability density function is

$$f(x) = 0.05e^{-0.05x},$$

with domain $[0, +\infty)$. An interesting and important question to ask is: What is the average length of time such an institution will last before failing? To answer this question, we use the following.

Mean or Expected Value

If X is a continuous random variable with probability density function f defined on an interval with (possibly infinite) endpoints a and b , then the **mean** or **expected value** of X is

$$E(X) = \int_a^b x f(x) dx.$$

$E(X)$ is also called the **average value** of X . It is what we expect to get if we take the average of many values of X obtained in experiments.

Example 1

Let X have probability density function given by

$$f(x) = 3x^2,$$

with domain $[0, 1]$. Find $E(X)$.

Solution

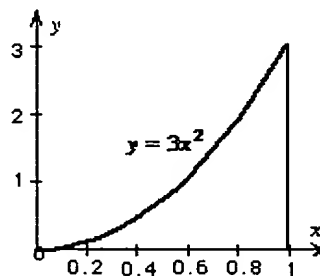
We have

$$\begin{aligned} E(X) &= \int_a^b x f(x) dx. \\ &= \int_0^1 (x)(3x^2) dx \\ &= \int_0^1 (3x^3) dx \\ &= [3x^4/4]_0^1 = 3/4. \end{aligned}$$

Thus, the expected value of X is $3/4$.

Before We Go On ...

This reflects the fact that X is more likely to take on values in the right part of the interval $[0, 1]$ than the left part. The figure shows this probability density function.



We shall explain shortly why $E(X)$ is given by the integral formula.

Example 2 *Failing S&Ls*

Given that troubled S&Ls are failing continuously at a rate of 5% per year, how long will the average troubled S&L last?

Solution

If X is the number of years that a given S&L will last, we know that its probability density function is $f(x) = 0.05e^{-0.05x}$. To answer the question we compute $E(X)$.

$$\begin{aligned}
 E(X) &= \int_a^b x f(x) dx. \\
 &= \int_0^{+\infty} (0.05xe^{-0.05x}) dx \\
 &= \lim_{M \rightarrow +\infty} \int_0^M (0.05xe^{-0.05x}) dx
 \end{aligned}$$

Using integration by parts, we get

$$E(X) = \lim_{M \rightarrow +\infty} -0.05 [e^{-0.05x}(20x + 400)]_0^M = (0.05)(400) = 20.$$

Thus, the expected lifespan of a troubled S&L is 20 years.

Before We Go On ...

Notice that the answer, 20, is the reciprocal of the failure rate 0.05. This is true in general: if $f(x) = ae^{-ax}$, then $E(X) = 1/a$.

Question

Why is $E(X)$ given by that integral formula?

Answer

Suppose for simplicity that the domain of f is a finite interval $[a, b]$. Break up the interval into n subintervals $[x_{k-1}, x_k]$, each of length Δx , as we did for Riemann sums. Now, the probability of seeing a value of X in $[x_{k-1}, x_k]$ is approximately $f(x_k) \Delta x$ (the approximate area under the graph of f over $[x_{k-1}, x_k]$). Think of this as the fraction of times we expect to see values of X in this range. These values, all close to x_k , then contribute approximately $x_k f(x_k) \Delta x$ to the average, if we average together many observations of X . Adding together all of these contributions, we get

$$E(X) \approx \sum_k x_k f(x_k) \Delta x$$

Now these approximations get better as $n \rightarrow \infty$, and we notice that the sum above is a Riemann sum converging to

$$E(X) = \int_a^b x f(x) dx,$$

which is the formula we have been using.

Question

What are the expected values of the standard distributions we discussed in the previous section?

Answer

Let's compute them one by one.

Mean of a Uniform Distribution

If X is uniformly distributed on $[a, b]$, then

$$E(X) = (a + b)/2.$$

This is not surprising, if you think about it for a minute. We'll leave the actual computation as one of the exercises.

Mean of an Exponential Distribution

If X has the exponential distribution function $f(x) = ae^{-ax}$, then


$$E(X) = 1/a.$$

We saw how to compute this in Example 2.

Mean of a Normal Distribution

If X is normally distributed with parameters μ and σ , then

$$E(X) = \mu.$$

This is why we called μ the mean, but we ought to do the calculation. Click on the footnote marker to see it. 

Mean of a Beta Distribution

If X has the beta distribution function $f(x) = (\beta + 1)(\beta + 2)x^\beta(1 - x)$, then

$$E(X) = (\beta + 1)/(\beta + 3)$$

Again, we shall leave this as an exercise.

Example 3 Downsizing in the Utilities Industry

A utilities industry consultant predicts a cutback in the Canadian Utilities industry during 2000-2005 by a percentage specified by a beta distribution with $\beta = 0.25$. What is the expected size of the cutback by Ontario Hydro? ☹

Solution

Since $\beta = 0.25$,

$$E(X) = (\beta + 1)/(\beta + 3) = 1.25/3.25 \approx 0.38.$$

Therefore, we can expect about a 38% cutback by Ontario Hydro.

Before We Go On ...

What $E(X)$ really tells us is that the *average* downsizing of many utilities will be 38%. Some will cut back more, and some will cut back less.

There is a generalization of the mean that we shall use below. If X is a random variable on the interval (a, b) and probability density function f , and if g is any function defined on that interval, then we can define the **expected value of g** to be

$$E(g(X)) = \int_a^b g(x)f(x) dx.$$

Thus, in particular, the mean is just the expected value of the function $g(x) = x$. We can interpret this as the average we expect if we compute $g(X)$ for many experimental values of X .

Variance and Standard Deviation

Statisticians use the variance and standard deviation of a continuous random variable X as a way of measuring its dispersion, or the degree to which it "scattered." The definitions are as follows.

Variance and Standard Deviation

Let X be a continuous random variable with density function f defined on the interval (a, b) , and let $\mu = E(X)$ be the mean of X . Then the **variance** of X is given by

$$\text{Var}(X) = E((X-\mu)^2) = \int_a^b (x-\mu)^2 f(x) dx.$$

The **standard deviation** of X is the square root of the variance,

$$\sigma(X) = (\text{Var}(X))^{0.5}$$

Notes

(1) In order to calculate the variance and standard deviation, we need first to calculate the mean.

(2) $\text{Var}(X)$ is the expected value of the function $(x-\mu)^2$, which measures the square of the distance of X from its mean. It is for this reason that $\text{Var}(X)$ is sometimes called the *mean square deviation*, and $\sigma(X)$ is called the *root mean square deviation*. $\text{Var}(X)$ will be larger if X tends to wander far away from its mean, and smaller if the values of X tend to cluster near its mean.

(3) The reason we take the square root in the definition of $\sigma(X)$ is that $\text{Var}(X)$ is the expected value of the *square* of the deviation from the mean, and thus is measured in square units. Its square root $\sigma(X)$ therefore gives us a measure in ordinary units.

Question

What are the variances and standard deviations of the standard distributions we discussed in the previous section?

Answer

Let's compute them one by one. We'll leave the actual computations (or special cases) for the exercises.

Variance and Standard Deviation of a Uniform Distribution

If X is uniformly distributed on $[a, b]$, then

$$\text{Var}(X) = (b-a)^2/12$$

and

$$\sigma(X) = (b-a)/12^{0.5}$$

Variance and Standard Deviation of an Exponential Distribution

If X has the exponential distribution function $f(x) = ae^{-ax}$, then

$$\text{Var}(X) = 1/a^2$$

and

$$\sigma(X) = 1/a.$$

Variance and Standard Deviation of a Normal Distribution

If X is normally distributed with parameters μ and σ , then

$$\text{Var}(X) = \sigma^2$$

and

$$\sigma(X) = \sigma.$$

(This is what you might have expected!)

Variance and Standard Deviation of a Beta Distribution

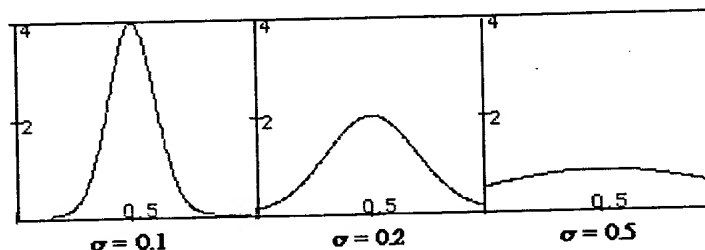
If X has the beta distribution function $f(x) = (\beta+1)(\beta+2)x^\beta(1-x)$, then

$$\text{Var}(X) = \frac{2(\beta+1)}{(\beta+4)(\beta+3)^2}$$

and

$$\sigma(X) = \sqrt{\frac{2(\beta+1)}{(\beta+4)(\beta+3)^2}}.$$

You can see the significance of the standard deviation quite clearly in the normal distribution. As we mentioned in the previous section, σ is the distance from the maximum at μ to the points of inflection at $\mu - \sigma$ and $\mu + \sigma$. The larger σ is, the wider the bell. The following shows three normal distributions with three different standard deviations (all with $\mu = 0.5$).



Again, a small standard deviation means that the values of X will be close to the mean with high probability, while a large standard deviation means that the values may wander far away with high probability.

Median

The *median income* in the U.S. is the income M such that half the population earn incomes $\leq M$ (so the other half earn incomes $> M$). In terms of probability, we can think of income as a random variable X . Then the probability that $X \leq M$ is $1/2$, and the probability that $X \geq M$ is also $1/2$.

Median

Let X be a continuous random variable. The **median** of X is the number M such that

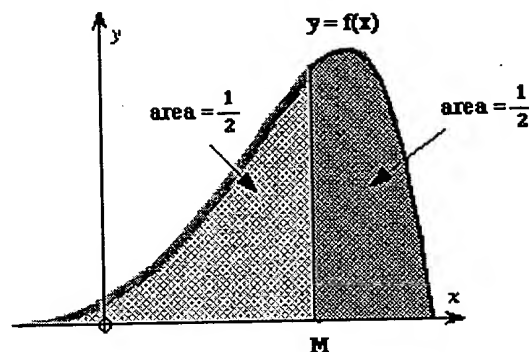
$$P(X \leq M) = 1/2.$$

Then $P(M \leq X) = 1/2$ also.

If f is the probability density function for X and f is defined on (a, b) , then we can calculate M by solving the equation

$$P(a \leq X \leq M) = \int_a^M f(x) \, dx = 1/2.$$

for M . Graphically, the vertical line $x = M$ divides the total area under the graph of f into two equal parts. (See the figure).



Question

What is the difference between the median and the mean?

Answer

Roughly speaking, the median divides the area under the distribution curve into two equal parts, while the mean is the value of X at which the graph would *balance*. If a probability curve has as much area to the left of the mean as to the right, then the mean is equal to the median. This is true of uniform and normal distributions, which are *symmetric* about their means. On the other hand, the medians and means are different for the exponential distributions and most of the beta distributions, because their areas are not distributed symmetrically.

Example 4 Lines at the Post Office

The time in minutes between individuals joining the line at an Ottawa Post Office is a random variable with the exponential distribution

$$f(x) = 2e^{-2x}, \quad (x \geq 0).$$

Find the mean and median time between individuals joining the line and interpret the answers.

Solution

The expected value for an exponential distribution $f(x) = ae^{-ax}$ is $1/a$. Here, $a = 2$, so $E(X) = 1/2$. We interpret this to mean that, on average, a new person will join the line every half a minute, or 30 seconds. For the median, we must solve

$$\int_0^M f(x) dx = 1/2.$$

That is,

$$\int_0^M (2 e^{-2x}) dx = 1/2.$$

Evaluating the integral gives

$$- [e^{-2x}]_0^M = 1/2,$$

or

$$1 - e^{-2M} = 1/2$$

so

$$e^{-2M} = 1/2,$$

or

$$-2M = \ln(1/2) = -\ln 2.$$

Thus,

$$M = (\ln 2)/2 \approx 0.3466 \text{ minutes.}$$

This means that half the people get in line less than 0.3466 minutes (about 21 seconds) after the previous person, while half arrive more than 0.3466 minutes later. The mean time for a new person to arrive in line is larger than this because there are some occasional long waits between people, and these pull the average up.

Sometimes we cannot solve the equation $\int_0^M f(x) dx = 1/2$ for M analytically, as the next example shows.

Example 5

Find the median of the random variable with beta density function for $\beta = 4$.

Solution

Here,

$$\begin{aligned} f(x) &= (\beta+1)(\beta+2)x^\beta(1-x) \\ &= 30x^4(1-x). \end{aligned}$$

Thus we must solve

$$\int_0^M (30x^4(1-x)) dx = 1/2.$$

That is,

$$30 \int_0^M (x^4 - x^5) dx = 1/2.$$

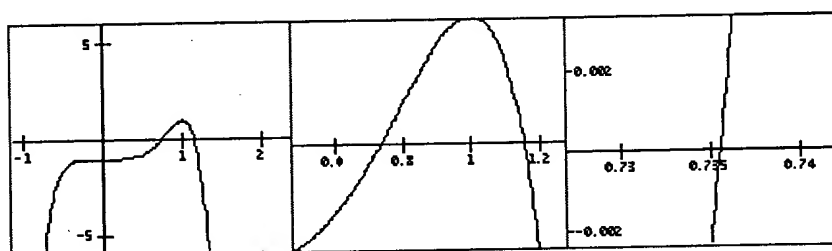
So

$$30[M^5/5 - M^6/6] = 1/2$$

or, multiplying through and clearing denominators,

$$12M^5 - 10M^6 - 1 = 0.$$

This is a degree six polynomial equation that has no easy factorization. Since there is no general analytical method for obtaining the solution, the only method we can use is numerical. The figure shows three successive views of a graphing calculator plot of $Y = 12X^5 - 10X^6 - 1$, obtained by zooming in towards one of the zeros.



We are interested only in the zero that occurs between 0 and 1 (why?), and find that $M \approx 0.735$ to within ± 0.001 .

Before We Go On ...

Question

This method required us first to calculate $\int_a^M f(x) dx$ analytically. What if even this is impossible to do?

Answer

We can solve the equation $\int_a^M f(x) dx = 1/2$ graphically by having the calculator compute and graph this function of M by numerical integration. For example, to redo the above example on the TI-83 or compatible models, enter


$$Y_1 = \text{fnInt}(30T^4(1-T), T, 0, X) - 0.5$$


which corresponds to

$$y = \int_0^x (30t^4(1-t)) dt - 1/2,$$


a function of x . Since the median of M is the solution obtained by setting $y = 0$, we can obtain the answer by plotting Y_1 and finding its x -intercept. The plot should be identical to the one we obtained above (why?).

2. Probability**Density Functions:**

 **Uniform,
Exponential, Normal,
and Beta**

 **Section 3
Exercises**

**4. You're the
Expert -
Creating a
Family Trust**

 **Calculus and
Probability
Main Page**

 **"Real
World"
Page**

We would welcome comments and suggestions for improving this resource.

Mail us at:

 Stefan Waner (matszw@hofstra.edu)  Steven R. Costenoble (matsrc@hofstra.edu)

Last Updated: *September, 1996*

Copyright © 1996 Stefan Waner and Steven R. Costenoble